y and Molecular Biology, Inc.

Vol. 265, No. 16, Issue of June 5, pp. 9153-9158, 1990
Printed in U.S.A.

The Complete S quenc of Botulinum Neurotoxin Type A and Comparison with Oth r Clostridial Neurotoxins*

(Received for publication, November 1, 1989)

Thomas Binz‡§, Hisao Kurazono‡¶, Micaela Wille‡, Jürgen Frevert‖, Karel Wernars**, and Heiner Niemann‡ ‡‡

*From the ‡Institut für Medizinische Virologie der Justus-Liebig-Universität, Giessen D-6300, Federal Republic of Germany, the ‖Battelle-Institut, Frankfurt D-6000, Federal Republic of Germany, and the **Rijksinstituut voor Volksgezondheid en Milieuhygiene, Bilthoven 3720 BA, The Netherlands*

The seven serologically different botulinum neurotoxins are highly potent protein toxins that inhibit n urotransmitter release from peripheral cholinergic synapses. The activated toxins consist of the toxifying A-subunits ($M_r \approx 50,000$) linked by a disulfide bond to the receptor-binding BC-subunits ($M_r \approx 100,000$). We have established the complete sequence of botulinum n urotoxin type A (BoNT/A; 1,296 amino acid residues, $M_r = 149,425$) and a partial sequence of botulinum neurotoxin type E (273 amino acid residues) as deduced from the corresponding nucleotide sequences of the chromosomally located structural genes. The promoter of the BoNT/A gene is inactive in *Escherichia coli*. Primer extension experiments indicated that initiation of transcription of the BoNT/A gene occurred 118 nucleotides upstream from the ATG codon. A comparison of the protein sequence revealed an overall identity of 33.8% to that of tetanus toxin. No significant similarity to other known proteins including ADP-ribosylating toxins could be detected. Three of the six histidine residues of the A-subunit of BoNT/A were found in the peptide sequence $H^{223}ELIHXXH^{230}$ within a domain of predicted α-helical secondary structure. This motif is also found in similar positions of the A-subunits of tetanus toxin and BoNT/E.

The Gram-positive spore-forming bacterium *Clostridium botulinum* produces several highly toxic types of botulinum neurotoxins (BoNTs)[1] designated A, B, C1, D1, E, F, and G. Today, the number of fatal cases of botulism is fortunately small due to improved hygienic precautions during production of food. In the past, however, predominantly BoNT/A and BoNT/E, the latter mostly found in decayed seafood, were identified as the agents causing life-threatening food poisoning of man (for a review, see Tacket and Rogawski, 1989). BoNTs resemble, in regard to biosynthesis and molecular architecture, tetanus toxin, whose sequence has already been determined (Eisel *et al.*, 1986; Fairweather and Lyness, 1986). The toxins are produced as single-chain polypeptides ($M_r \approx$ 150,000) which are, with the exception of BoNT/E, proteolytically processed into two subunits upon release from the organism. The two subunits, designated A- and BC-fragments (corresponding to the light and heavy chains), remain covalently associated by a disulfide bond. Toxification by tetanus toxin and BoNTs is initiated by binding of the toxin molecules to specific ganglioside receptors, followed by receptor-mediated endocytosis (Black and Dolly, 1986a, 1986b). In contrast to tetanus toxin, which acts preferentially after retrograde axonal transport by blocking the release of inhibitory neurotransmitters from central synapses, BoNTs evoke their toxic effects in the periphery by poisoning cholinergic synapses (Simpson, 1989). Recent studies based on the administration of isolated A-fragments from tetanus toxin or from BoNT/A to permeabilized bovine chromaffin cells have shown that the A-fragments alone are sufficient to inhibit exocytosis (Ahnert-Hilger *et al.*, 1989; Bittner *et al.*, 1989). The molecular mechanisms, however, that underlie toxification of central or peripheral synapses through clostridial neurotoxins still remain completely unclear. The characterization of BoNT genes and the combined *in vitro* transcription/translations of subfragments thereof could provide useful tools for identifying cellular components involved in late stages of transmitter release. Furthermore, the application of specific DNA probes could be of advantage in the quality control of food production. Finally, the genetic development of nontoxic derivatives should allow the production of safer vaccines.

In this study, we present for the first time a complete sequence of a botulinum neurotoxin, BoNT/A, as well as part of the sequence of BoNT/E. We compare these sequences with those of tetanus toxin and partial amino acid sequences of other botulinum toxins. We show that a predicted α-helical domain within the A-subunit containing 3 histidine residues is conserved in tetanus toxin and BoNT/A and BoNT/E. In addition, we have mapped th transcription start site of the BoNT/A gene. Our data suggest that BoNT/A is translated from monocistronic mRNA.

MATERIALS AND METHODS

Biosafety Regulations—Personnel involved in the isolation and cloning of DNA from *C. botulinum* were immunized with a pentameric

* This work was supported in part by Grant Nie 175/5-2 from the Deutsche Forschungsgemeinschaft (to H. N.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Dedicated to Professor Dr. S. Stirm, Giessen.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M30196.

§ Supported by a doctoral fellowship from the University of Giessen. This paper was submitted in order to fulfill the requirements for a doctorate of natural sciences.

¶ Supported by an Alexander von Humboldt fellowship.

‡‡ To whom correspondence and reprint requests should be addressed: Inst. for Microbiology, Federal Research Center for Virus Diseases of Animals, P. O. Box 1149, Paul-Ehrlich-Str. 28, Tübingen D-7400, Federal Republic of Germany.

[1] The abbreviations used are: BoNTs, botulinum neurotoxins; BoNT/A, botulinum neurotoxin type A; BoNT/E, botulinum neurotoxin type E; DEPC, diethyl pyrocarbonate; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; kb, kilobase pair(s).

vaccine consisting of formaldehyde-treated BoNT/A through BoNT/E. Only DNA fragments smaller than 2.0 kb (1.0 kb for BoNT/E-specific DNA) were cloned under L3B1 containment following the protocols of Maniatis *et al.* (1982). Toxicity assays were performed in mice using 200 μl of a bacterial lysate obtained from 3×10^6 bacteria (Boroff and Fleck, 1966). None of the recombinant bacterial clones exhibited toxicity. Recombinant DNA was isolated from JM83 and sequenced directly using the chain termination method (Sanger *et al.*, 1977).

Bacterial DNA and Plasmids—For the isolation of chromosomal DNA from *C. botulinum* (subtype A, strain 62A), overnight precultures were prepared in 10 ml of brain-heart infusion broth and used to inoculate a 190-ml culture. The culture was incubated at 37 °C for about 8 h to $OD_{540\,nm} = 0.3$. *C. botulinum* type E (strain Beluga) was grown for 24 h in 500 ml of broth containing 2% (w/v) proteose peptone, 0.5% (w/v) glucose, and 0.1% (w/v) cysteine. The cells were centrifuged at 4 °C and $5{,}000 \times g$ for 15 min and resuspended in 8 ml of TE buffer (10 mM Tris/HCl, 25 mM EDTA, pH 8.0) containing 2 mg/ml lysozyme. The suspension was incubated for 1 h at 37 °C. At this time, the EDTA concentration was raised to 100 mM, and Sarkosyl (1% final concentration) and proteinase K (0.1 mg/ml final concentration) were added. The mixture was extracted twice with phenol and once more with chloroform prior to dialysis against three changes of 1,000 ml of TE buffer. The mixture was treated for 2 h at 37 °C with RNase A (0.1 mg/ml final concentration), extracted with phenol:chloroform (1:1), and precipitated with ethanol. Plasmids pSP65 and pSP64 (Krieg and Melton, 1984) were used to establish BoNT/A-specific DNA libraries and for *in vitro* transcription. *Escherichia coli* strain HB101 was used for the amplification of plasmid DNA. BoNT/E-specific libraries were constructed in pUC19 (Yanisch-Perron *et al.*, 1985) using *E. coli* strain DH5α.

Enzymes and Nucleotides—Restriction endonucleases, T4 DNA ligase, and T4 polynucleotide kinase were from Boehringer Mannheim. Nucleotides and T7 polymerase were from P-L Biochemicals. Exonuclease III, S1 nuclease, and Klenow polymerase were purchased from Promega Biotec. A nick translation system and radiolabeled [α-^{32}P]dCTP (>3000 Ci/mmol), [γ-^{32}P]ATP (>3000 Ci/mmol), and ^{35}S-dCTP (>800 Ci/mmol) were obtained from Amersham Corp.

Hybridization Conditions—Oligonucleotides were synthesized by the phosphoramidite method using an Applied Biosystems Model 380A automated synthesizer. To identify BoNT/A-specific sequences, a single oligonucleotide (5′-AAACAATTTAATTATAAAGATCC-3′) was used. This sequence corresponds to Lys^6 to Pro^{12} of a partial amino acid sequence published previously (Schmidt *et al.*, 1984). BoNT/E-specific DNA was screened with a mixture of 29-mer oligonucleotides (5′-TTTAA(C/T)TATAA(C/T)GATCCTGTTAATGA(C/T)CG-3′) whose sequences corresponded to Phe^7 to Arg^{16} from the partial amino acid sequence reported by Schmidt *et al.* (1985). GeneScreen *Plus* filters (Du Pont-New England Nuclear) were prehybridized in 6 × SSC, 10 × Denhardt's solution, and 100 μg of sheared denatured herring sperm DNA/ml at 43 °C. Hybridizations were performed with 10^7 cpm of the oligonucleotide probe for 24 h at 43 °C. The filters containing BoNT/A-specific DNA fragments were washed twice for 10 min at 43 °C in 6 × SSC, 0.1% SDS. Washing of BoNT/E DNA-containing filters was performed at 45 °C in 2 × SSC, 1% SDS, 0.1% sodium pyrophosphate. Filters were exposed overnight to Fuji RX films. Southern analyses involving nick-translated probes were performed as described (Maniatis *et al.*, 1982).

Combined in Vitro Transcription/Translation of BoNT/A-specific DNA Fragments—Plasmid pKN25, harboring the 1.5-kb *Ssp*I fragment of the BoNT/A gene, was purified by two successive centrifugation steps on CsCl gradients or by a single passage over a QIAGEN column (Diagen, Düsseldorf, Federal Republic of Germany). The DNA was linearized downstream from the coding region by digestion with *Xba*I. Transcriptions and translations in a reticulocyte lysate (Amersham Corp.) were performed as described (Mayer *et al.*, 1988), using [^{35}S]methionine (Du Pont-New England Nuclear; 1 mCi/ml final concentration) as radiolabeled amino acid. Translation products were analyzed by SDS-PAGE on 12% gels.

Mapping Initiation Site for Transcription—Total RNA was isolated from *C. botulinum* as described (Ausubel *et al.*, 1987). For primer extension experiments, a heptadecamer oligonucleotide whose sequence was complementary to nucleotides −10 to +7 (Fig. 2) was 5′-labeled with [γ-^{32}P]ATP and applied in the primer extension reaction as described (Bensi *et al.*, 1986) and in the sequencing reaction using pKN13 DNA as a template.

Antibodies—Polyclonal and monoclonal antibodies against formaldehyde-treated BoNT/A were kindly provided by Dr. S. Kozaki and Dr. G. Sakaguchi (Osaka Prefecture University, Osaka, Japan). The production of polyclonal antibodies has been described previously (Kozaki and Sakaguchi, 1975). Toxicity assays were performed in mice according to Boroff and Fleck (1966).

Modification of BoNT/A with DEPC—Carboxyethylation reactions of purified BoNT/A (2.6 nmol) were performed with a 1000-fold molar excess of DEPC (ethoxyformic anhydride; Sigma) in 1 ml of 100 mM sodium phosphate, pH 6.0, essentially as described by DasGupta and Rasmussen (1984). The absorption at 240 nm was monitored using the same amount of toxin in buffer as a blank. DEPC was added from a stock solution in ethanol. The blank was treated with ethanol only. For calculation of the concentration of *N*-carboxyhistidine, an extinction coefficient of 3200 $M^{-1}\cdot cm^{-1}$ was used. Reversal of the DEPC modification was attempted with DEPC toxoid obtained after a 1-h treatment. The procedure for incubation with $NH_2OH \cdot HCl$ was as described previously (DasGupta and Rasmussen, 1984). In our hands, however, this treatment failed to restore toxicity.

RESULTS AND DISCUSSION

Previous studies on the structural gene of tetanus toxin revealed the strong bias of clostridia toward (A + T)-rich codons (Eisel *et al.*, 1986). For this reason, we used oligonucleotide probes reflecting this specific codon usage. The BoNT/A probe specifically recognized chromosomal DNA fragments of 3.7, 1.40, 1.50, and 0.45 kb as obtained by digestion with endonucleases *Eco*RI, *Eco*RI/*Pvu*II, *Ssp*I, and *Hin*dIII, respectively (Fig. 1*A*). DNA migrating in the region of the 1.50-kb *Ssp*I band was electroeluted from the agarose gel and cloned under L3 biosafety conditions into the *Sma*I site of pSP65. pKN2 and pKN25 contained the desired fragment in opposite orientations. Direct sequencing of pKN2 and pKN25 and exonuclease III-truncated variants of both clones revealed that the inserted 1498-base pair *Ssp*I fragment encoded the entire A-subunit of BoNT/A. Antisense RNA was transcribed from pKN2 and used to screen additional DNA libraries. The entire sequence of the BoNT/A gene was established from five overlapping clones (see Fig. 1*B*) and deleted variants thereof. To further verity the identity of the cloned *Ssp*I fragment, we performed combined *in vitro* transcription/translations using purified pKN25 DNA. For this purpose, no modifications of pKN25 were required since sequencing of the inserted *Ssp*I fragment had revealed that the authentic ATG codon for initiation of translation of BoNT/A was indeed the first ATG codon 59 nucleotides downstr am from the SP6 promoter. The products obtained in the reticulocyte lysate were analyzed by SDS-PAGE (Fig. 1*C*). The majority of the radiolabel was incorporated into a molecular species of about 56 kDa (*lane 1*), as would be expected from the coding capacity of the *Ssp*I fragment. The polypeptide was specifically precipitated by an A-subunit-specific monoclonal antibody (*lanes 3* and *4*) and by a polyclonal rabbit serum raised against complete BoNT/A toxoid (*lane 5*).

The 5′-radiolabeled BoNT/E-specific oligonucleotide hybridized to a single 984-base pair *Eco*RI fragment from chromosomal type E-specific DNA (data not shown). The localization of this fragment within the BoNT/E gene is shown in Fig. 1*B*. The fragment was cloned into pUC19 and sequenced (Fig. 2*B*).

Fig. 2*A* shows a continuous stretch of 4,836 nucleotides derived from chromosomal DNA of *C. botulinum* type A containing a single open reading frame of 3,888 nucleotides. The corresponding protein (1,296 amino acid residues) has a molecular mass of 149,425 Da. The A + T c ntent in the 5′-noncoding regions of the BoNT/A and BoNT/E genes is 80.4 and 80.3%, respectively, and thus higher than in the corresponding coding regions where a total of 73.6 and 72.1% A + T is found. The codon usage of the BoNT/A gene was closely related to that of the tetanus toxin gene. A total of 63.4% (63.1% in tetanus toxin) A + T was found at positi n 1, 70.7%

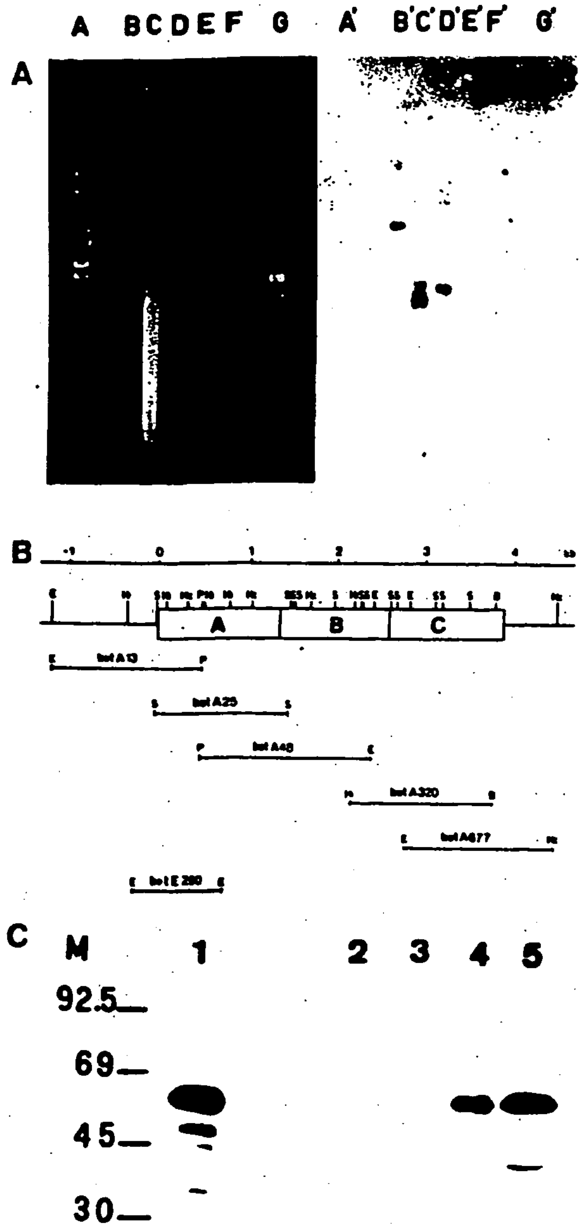


FIG. 1. **Identification and organization of structural gene encoding BoNT/A and BoNT/E.** *A*, Southern analyses of chromosomal DNA from *C. botulinum* type A after digestion with *Eco*RI (*lane B*), *Ssp*I (*lane C*), *Eco*RI/*Pvu*II (*lane D*), *Hin*dIII (*lane E*), or *Xba*I (*lane F*). *Lanes A* and *G* show λDNA cleaved with *Hin*dIII and *Eco*RI/*Hin*dIII, respectively. *Lanes A′–G′* show the results after hybridization with the 5′-labeled oligonucleotide. *B*, the coding region for BoNT/A containing the A-, B-, and C-subfragments is indicated by the *open boxes*. *Horizontal bars* and designations indicate individual clones used for sequencing. *B*, *Bgl*II; *E*, *Eco*RI; *Hc*, *Hin*cII; *Hi*, *Hin*dIII; *S*, *Ssp*I; *P*, *Pvu*II. The localization of the 984-base pair *Eco*RI fragment from BoNT/E is included at the bottom. *C*, *in vitro* transcription/translation of pKN25. The 1.5-kb *Ssp*I insert of pKN25 encodes a polypeptide of 56,364 Da that contains the entire A-subunit. 5′-Capped mRNA was obtained with SP6 RNA polymerase and translated in the presence of [^{35}S]methionine in a reticulocyte lysate. Products were analyzed by SDS-PAGE on a 12% gel (*lane 1*) or subjected to immunoprecipitation using preimmune serum (*lane 2*), monoclonal antibody E-2 at 10 μg of purified IgG/ml (*lane 3*) or at a 100 μg/ml final concentration (*lane 4*), or a polyclonal rabbit serum (*lane 5*).

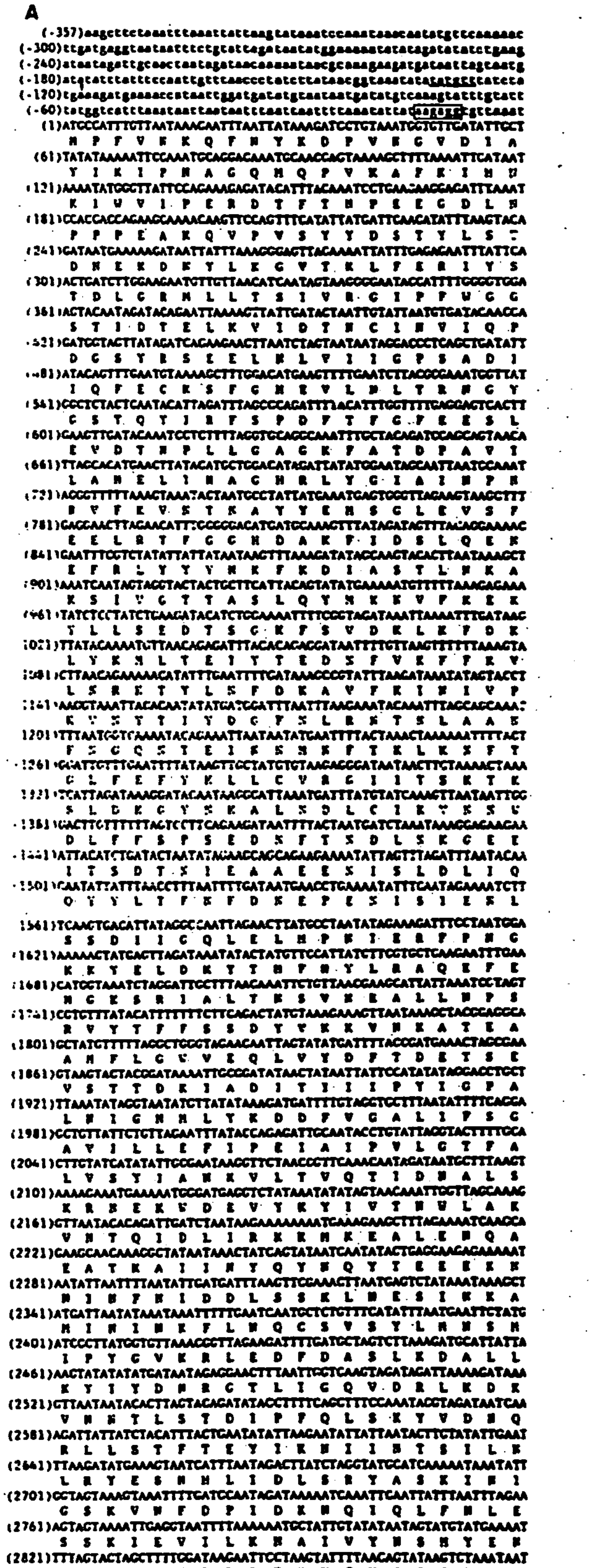


FIG. 2. **Nucleotide and amino acid sequences of BoNT/A gene (*A*) and of BoNT/E-specific *Eco*RI fragment (*B*).** The *arrowhead* marks the initiation site for transcription. The −10 region of the toxin promoter is *underlined*. Shine-Dalgarno sequences are *boxed*. *Divergent arrows* in the 3′-noncoding region of the BoNT/A gene indicate motifs with dyad symmetry. Numbers define the melting energy in kilocalories as calculated according to Tinoco *et al.* (1973). Encoded amino acids are in single-letter code.

```
(2881)GAATATACAATAATAAATTGTATGGAAAATAATTCAGGATGGAAAGTATCACTTAATTAT
       E  Y  T  I  I  N  C  M  E  N  N  S  G  W  K  V  S  L  N  Y
(2941)GGTGAAATAATCTGGACTTTACAGGATACTCAGGAAATAAAACAAAGAGTAGTTTTTAAA
       G  E  I  I  W  T  L  Q  D  T  Q  E  I  K  Q  R  V  V  F  K
(3001)TACAGTCAAATGATTAATATATCAGATTATATAAACAGATGGATTTTTGTAACTATCACT
       Y  S  Q  M  I  N  I  S  D  Y  I  N  R  W  I  F  V  T  I  T
(3061)AATAATAGATTAAATAACTCTAAAATTTATATAAATGGAAGATTAATAGATCAAAAACCA
       N  N  R  L  N  N  S  K  I  Y  I  N  G  R  L  I  D  Q  K  P
(3121)ATTTCAAATTTAGGTAATATTCATGCTAGTAATAATATAATGTTTAAATTAGATGGTTGT
       I  S  N  L  G  N  I  H  A  S  N  N  I  M  F  K  L  D  G  C
(3181)AGAGATACACATAGATATATTTGGATAAAATATTTTAATCTTTTTGATAAGGAATTAAAT
       R  D  T  H  R  Y  I  W  I  K  Y  F  N  L  F  D  K  E  L  N
(3241)GAAAAAGAAATCAAAGATTTATATGATAATCAATCAAATTCAGGTATTTTAAAAGACTTT
       E  K  E  I  K  D  L  Y  D  N  Q  S  N  S  G  I  L  K  D  F
(3301)TGGGGTGATTATTTACAATATGATAAACCATACTATATGTTAAATTTATATGATCCAAAT
       W  G  D  Y  L  Q  Y  D  K  P  Y  Y  M  L  N  L  Y  D  P  N
(3361)AAATATGTCGATGTAAATAATGTAGGTATTAGAGGTTATATGTATCTTAAAGGGCCTAGA
       K  Y  V  D  V  N  N  V  G  I  R  G  Y  M  Y  L  K  G  P  R
(3421)GGTAGCGTAATGACTACAAACATTTATTTAAATTCAAGTTTGTATAGGGGGACAAAATTT
       G  S  V  M  T  T  N  I  Y  L  N  S  S  L  Y  R  G  T  K  F
(3481)ATTATAAAAAAATATGCTTCTGGAAATAAAGATAATATTGTTAGAAATAATGATCGTGTA
       I  I  K  K  Y  A  S  G  N  K  D  N  I  V  R  N  N  D  R  V
(3541)TATATTAATGTAGTAGTTAAAAATAAAGAATATAGGTTAGCTACTAATGCATCACAGGCA
       Y  I  N  V  V  V  K  N  K  E  Y  R  L  A  T  N  A  S  Q  A
(3601)GGCGTAGAAAAAATACTAAGTGCATTAGAAATACCTGATGTAGGAAATCTAAGTCAAGTA
       G  V  E  K  I  L  S  A  L  E  I  P  D  V  G  N  L  S  Q  V
(3661)GTAGTAATGAAGTCAAAAAATGATCAAGGAATAACAAATAAATGCAAAATGAATTTACAA
       V  V  M  K  S  K  N  D  Q  G  I  T  N  K  C  K  M  N  L  Q
(3721)GATAATAATGGCAATGATATAGGCTTTATAGGATTTCATCAGTTTAATAATATAGCTAAA
       D  N  N  G  N  D  I  G  F  I  G  F  H  Q  F  N  N  I  A  K
(3781)CTAGTAGCAAGTAATTGGTATAATAGACAAATAGAAAGATCTAGTAGGACTTTGGGTTGC
       L  V  A  S  N  W  Y  N  R  Q  I  E  R  S  S  R  T  L  G  C
(3841)TCATGGGAATTTATTCCTGTAGATGATGGATGGGGAGAAAGGCCACT taattaatctca
       S  W  E  F  I  P  V  D  D  G  W  G  E  R  P  L  *
(3901)aactacatgagtctgtcaagaattttctgtaaacatccataaaattttaaattaatat
(3961)gtttaagaataactagatatgagtattgtttgaactgcccctgtcaagtagacaggtaaa
                                                <----- -8.8----->
(4021)aaaataaaaattaagatactatggtctgatttcgatattctatcggagtcagaccttttta
                                  <-------------- -17.0 -------------->
(4081)actttttcttgtatcctttttgtattgtaaaactctatgtattcatcaattgcaagttcca
(4141)attagtcaaattatgaaactttctaagataatacatttctgattttatattttccaaa
(4201)atcctttccataggaccattatcaatacatctaccaactcgagacatactttgagttgtgt
(4261)ctatcccattaagtttattcttgaaagatttacttgtatattgaaaccgctatacatg
(4321)gaaaaagtggactagcatcaggattggaggtaactgctttatcaaaggtttcaaagacaa
(4381)ggacgttgttatttgattttccaagtacataggaataatgctattatcatgcaaatcaa
(4421)gtatttcactcaagtacgcctttgtttcgtctgttaac
```

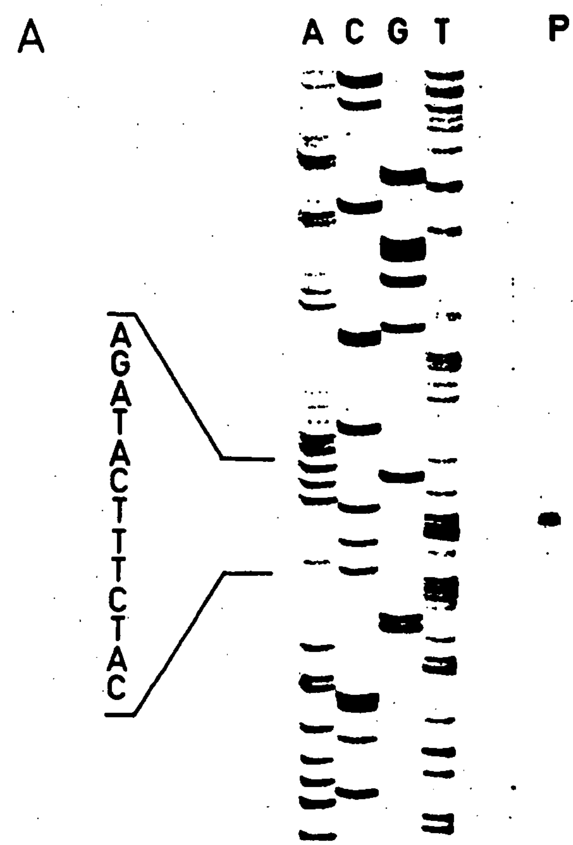
B

```
      (-228)gaattcaagtagtagataataaaaataatgccacagatttttatttatt
(-180)aataatgatatatttatctctaactgtttaactttaacttataacaatgtaaatatatat
(-120)ttgtctataaaaaatcaagattacaattgggttatatgtgatcttaatcatgatatacca
 (-60)aaaaagtcatatctatggatattaaaaaatatataaatttaaaattggaggagtgtgtat
   (1)ATGCCAAAAATTAATAGTTTTAATTATAATGATCCTGTTAATGATAGAACAATTTTATAT
       M  P  K  I  N  S  F  N  Y  N  D  P  V  N  D  R  T  I  L  Y
  (61)ATTAAACCAGGCGGTTGTCAAGAATTTTATAAATCATTTAATATTATGAAAAATATTTGG
       I  K  P  G  G  C  Q  E  F  Y  K  S  F  N  I  M  K  N  I  W
 (121)ATAATTCCAGAGAGAAATGTAATTGGTACAACCCCCCAAGATTTTCATCCGCCTACTTCA
       I  I  P  E  R  N  V  I  G  T  T  P  Q  D  F  H  P  P  T  S
 (181)TTAAAAAATGGAGATAGTAGTTATTATGACCCTAATTATTTACAAAGTGATGAAGAAAAG
       L  K  N  G  D  S  S  Y  Y  D  P  N  Y  L  Q  S  D  E  E  K
 (241)GATAGATTTTTAAAAATAGTCACAAAAATATTTAATAGAATAAATAATAATCTTTCAGGA
       D  R  F  L  K  I  V  T  K  I  F  N  R  I  N  N  N  L  S  G
 (301)GGGATTTTATTAGAAGAACTGTCAAAAGCTAATCCATATTTAGGGAATGATAATACTCCA
       G  I  L  L  E  E  L  S  K  A  N  P  Y  L  G  N  D  N  T  P
 (361)GATAATCAATTCCATATTGGTGATGCATCAGCAGTTGAGATTAAATTCTCAAATGGTAGC
       D  N  Q  F  H  I  G  D  A  S  A  V  E  I  K  F  S  N  G  S
 (421)CAAGACATACTATTACCTAATGTTATTATAATGGGAGCAGAGCCTGATTTATTTGAAACT
       Q  D  I  L  L  P  N  V  I  I  M  G  A  E  P  D  L  F  E  T
 (481)AACAGTTCCAATATTTCTCTAAGAAATAATTATATGCCAAGCAATCACCGTTTTGGATCA
       N  S  S  N  I  S  L  R  N  N  Y  M  P  S  N  H  R  F  G  S
 (541)ATAGCTATAGTAACATTCTCACCTGAATATTCTTTTAGATTTAATGATAATAGTATGAAT
       I  A  I  V  T  F  S  P  E  Y  S  F  R  F  N  D  N  S  M  N
 (601)GAATTTATTCAAGATCCTGCTCTTACATTAATGCATGAATTAATACATTCATTACATGGA
       E  F  I  Q  D  P  A  L  T  L  M  H  E  L  I  H  S  L  H  G
 (661)CTATATGGGGCTAAAGGGATTACTACAAAGTATACTATAACACAAAAACAAAATCCCCTA
       L  Y  G  A  K  G  I  T  T  K  Y  T  I  T  Q  K  Q  N  P  L
 (721)ATAACAAATATAAGAGGTACAAATATTGAAGAATTC
       I  T  N  I  R  G  T  N  I  E  E  F
```

FIG. 2—*continued*

(69.4%) at position 2, and 86.7% (83.7%) at position 3.

We have reported previously that the promoter controlling the tetanus toxin gene is located on plasmid pE88 immediately upstream from the coding region (Eisel *et al.*, 1986). Transcription was shown to start 127 nucleotides upstream from the ATG codon for translation (Niemann *et al.*, 1988). To determine whether the chromosomally located BoNT/A gene was part of an operon or whether it was transcribed into a monocistronic mRNA, we mapped the transcription start site by primer extension experiments using clostridial mRNA as a template. The results presented in Fig. 3*A* demonstrate that transcription begins 118 nucleotides upstream from the ATG start codon. Fig. 3*B* shows an alignment of the putative promoter regions of the BoNT/A and tetanus toxin promoters in comparison to a consensus sequence compiled from 29 different promoters from Gram-positive bacteria (Graves and Rabinowitz, 1986) and the lac promoter. In general, a precise allocation of the −35 and −10 regions is difficult within DNA sequences containing about 80% A + T. Therefore, the assignment of these regions in Fig. 3*B* is tentative only. The numeric −35 regions of the toxin promoters show almost no similarity to the Gram-positive consensus promoter. More closely related −35 motifs are indicated by the *asterisks.* These regions, however, would be separated from the corresponding Pribnow boxes by 22 and 21 nucleotides, a distance nly



B

```
bot a                 -35                                  -10    -5  +1
ATTTCCAATTGTTTAA  CCCTAT  CTTATAACGGTAAATATA  TATGTT  TATCTATGAAA
            ****  **

tet                   -35                                  -10    -5  +1
TAAAAATTTAAATTT  TCAGTT  TACAAAAAATAACCTGAT  TATGTT  ATATGTAATTG
            ***  ***  --  ----

Gram+               -35                                 -10    -5  +1
     Ta-AAAAA----  TTGACA  ---a--A---a-T-TG-  TATAAT  AAtAt----

lac                 -35                                 -10    -5  +1
     GGCACCCCAGGC  TTTACA  CTTTATGCTTCCGGCTCG  TATGTT  GTGTGGAA
```

FIG. 3. **Characterization of promoter of BoNT/A gene.** *A*, mapping of transcription start site. mRNA was isolated from *C. botulinum* type A. The oligonucleotide 5′-ATGGCATATTTAA-CACC-3′ complementary to nucleotides −10 to +7 of the BoNT/A gene was 5′-radiolabeled and used in a primer extension experiment using reverse transcriptase and for direct sequencing with T7 DNA polymerase. The products of both reactions were analyzed on the same gel. For the detection of the reverse transcribed product (*lane P*), however, the gel was exposed 3-fold longer (3 days). *B*, comparison of promoter regions. The putative −35 and −10 regions are listed for the BoNT/A (*bot a*) and tetanus toxin (*tet*) genes and compared with a consensus promoter sequence established from 29 different promoter sequences from Gram-positive bacteria (Graves and Rabinowitz, 1986) and the lac promoter. *Vertical arrows* indicate initiation sites for transcription. The promoter of the tetanus toxin gene has been characterized previously (Niemann *et al.*, 1988). For further details, see the text.

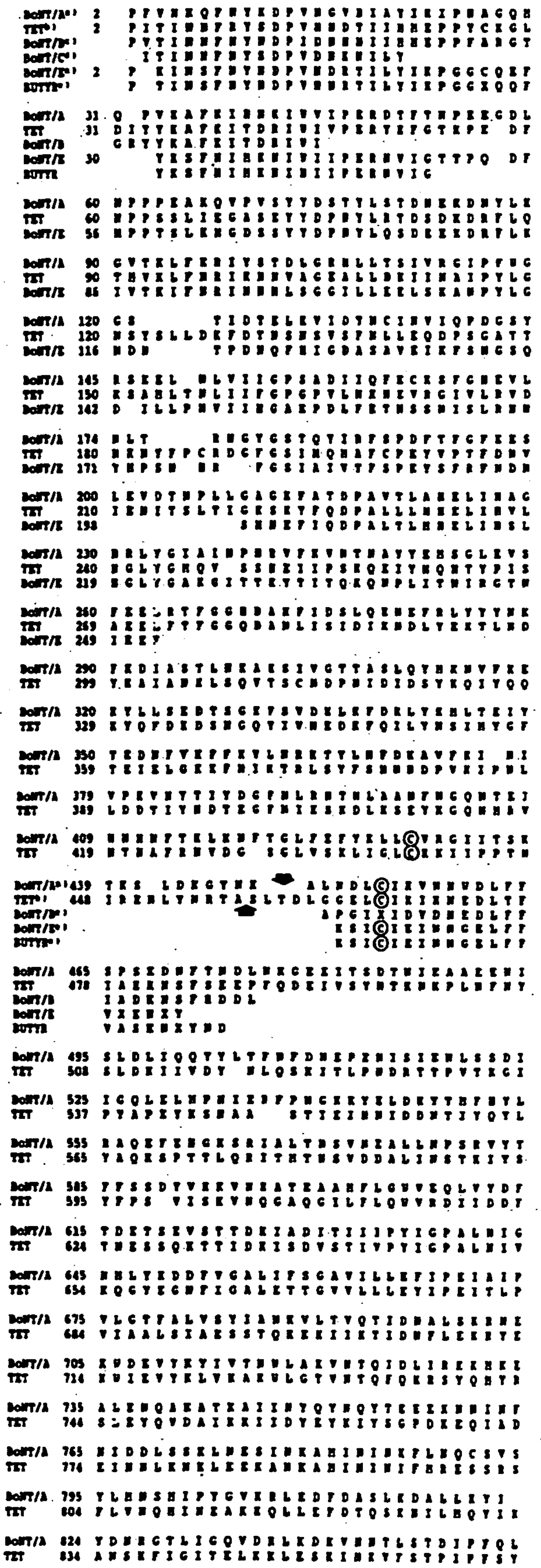


FIG. 4. Alignment of sequences of clostridial neurotoxins. Conserved cysteine residues are *circled*. Sequences were established in this paper (*a*) or were from Eisel *et al.* (1986) (*b*), DasGupta and Datta (1988) (*c*), Fujii *et al.* (1988) (*d*), and Gimenez *et al.* (1988) (*e*). TET, tetanus toxin; BUTYR, BoNT/E (from *Clostridium butyricum*, Gimenez *et al.* 1988).

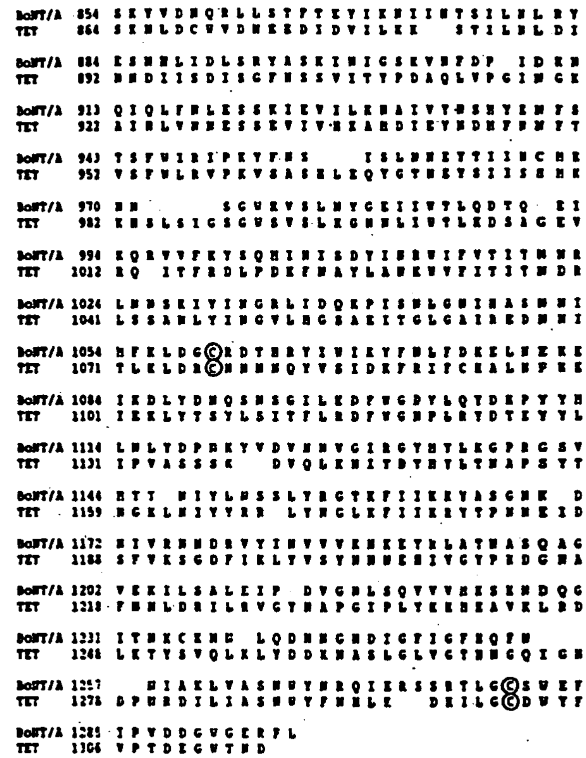


FIG. 4—*continued*

seldomly observed in Gram-positive promoters (Graves and Rabinowitz, 1986). The −10 regions of the BoNT/A and tetanus toxin promoters were identical to that of the lac promoter. However, in contrast to the tetanus toxin promoter that had been shown to be active in *E. coli* (Eisel *et al.*, 1986) and exhibit about one-third of the strength of the tac promoter (Niemann *et al.*, 1988), the promoter of the BoNT/A gene was apparently inactive in *E. coli* since neither toxin-specific peptides nor transcripts could be demonstrated in pKN13-transfected HB101 cells (data not shown). Interestingly, the −35 motif (5′-TTTACA-3′), characteristic of the lac promoter, was also found in the DNA sequence of the tetanus toxin gene (*underlined* in Fig. 3*B*). Site-directed mutagenesis of this latter motif will provide information on whether the observed similarities are indeed important for the activity of the tetanus toxin promoter in *E. coli*. Taking into account that the 3′-noncoding region of the BoNT/A gene contains two regions with dyad symmetry (*underlined* by *divergent arrows* in Fig. 2*A*) which are characterized by free energy values (Tinoco *et al.*, 1973), *dG*(25 °C), of up to −17 kcal and which therefore probably constitute transcriptional barriers, we postulate that BoNT/A, like tetanus toxin, is translated from monocistronic mRNA.

A comparison of the complete amino acid sequence of BoNT/A with that of tetanus toxin and the partial sequence deduced from the BoNT/E-specific *Eco*RI fragment is shown in Fig. 4. The amino acid composition of BoNT/A closely resembles constituent analyses performed on purified A- and BC-subunits (Sathyamoorthy and DasGupta, 1985). Furthermore, the sequence is in total agreement with: (i) a partial nucleotide sequence btained f r the NH_2-terminal region of the A-subunit (Betley *et al.*, 1989); (ii) the sequence of the NH_2 terminus of the BC-subunit (Sathyamoorthy *et al.*, 1986) as determined by Edman degradation; and (iii) an internal peptide sequence from the BC-subunit containing Ile[1148] to Leu[1217,2].

[2] B. R. DasGupta, personal communication.

The comparison of the two sequences newly established in this study with th sequence of tetanus toxin reveals several interesting features. (i) There is an overall 33.8% identity between BoNT/A and tetanus toxin at the amino acid level (30.8% within the A-subunit and 35.4% within the BC-subunit). Although these values increase to 40.5% for the A-subunit and 48.1% for the BC-subunit when conservative exchanges such as Arg to Lys, Glu to Asp, etc. are neglected, the identity is much smaller than originally anticipated (Eisel *et al.*, 1986). The identity between the BoNT/E-specific subfragment (containing 273 amino acid residues) and the corresponding domain of tetanus toxin amounts to 30.4% (41.4% excluding conserved exchanges) as compared to 38.8% (45.8%) identity between the same domains of BoNT/E and BoNT/A. The identity between tetanus toxin and BoNT/A in this region is 33.3% (42.1%).

(ii) From Thr^{623} to Phe^{787} of the B-fragment, the identity reaches 47.9% (61.8% including conservative exchanges). This domain of the tetanus toxin molecule has been shown to form channels in asolectin vesicles at a pH below 5.0 (Roa and Boquet, 1985) and one may speculate that this domain provides the machinery for translocation of the individual A-subunits into the cytosol.

(iii) The alignment of the sequences shown in Fig. 4 yields 4 conserved cysteine residues (involving Cys^{430}, Cys^{454}, Cys^{1060}, and Cys^{1280} of BoNT/A), suggesting that these residues may be involved in disulfide bridges. Indeed, the former 2 cysteine residues participate in the disulfide bond between the A- and BC-subunits. Recent data by Krieglstein *et al.* (1990) indicate that the internal disulfide bridge present in the C-fragment f tetanus toxin involves Cys^{1076} and Cys^{1092}. No such small loop structure could be formed in BoNT/A. At present, however, no data are available that would indicate whether the internal disulfide bridge present within the C-fragment of tetanus toxin is important for binding to specific ganglioside receptors. In addition, the corresponding domain of BoNT/A has not been analyzed for the presence of internal disulfide bonds.

(iv) The A-subunit of BoNT/A contains 6 histidine residues, 3 of which (His^{223}, His^{227}, and His^{230}) are conserved in tetanus toxin and BoNT/E. These histidine residues appear in a short peptide sequence which has a predicted α-helical structure according to the rules of Chou and Fasman (1978). Due to the specific localization of the 3 histidine residues within the helix, the imidazole side groups of the histidine residues should appear at the same face of the helix winding, allowing these side groups to interact with each other or with an as yet undefined cellular ligand. DEPC, a reagent specific for histidine, lysine, and serine residues, was shown to detoxify BoNT/E and, to a lesser extent, also BoNT/A (DasGupta and Rasmussen, 1984). In our experiments, treatment of BoNT/A with a 1000-fold molar excess of DEPC led to complete detoxification after the incorporation of 30 mol of reagent/mol of protein. Although the immunological properties of BoNT/A were not altered by such treatment, as assessed by Ouchterlony gel diffusion assays, toxicity could not be restored by subsequent treatment with hydroxylamine (data not shown). This indicates that other residues in addition to the histidines are irreversibly modified by DEPC. Site-directed mutagenesis of the conserved histidines within the α-helical sequences in the three neurotoxins should reveal whether this motif contributes to the biol gical function of the neurotoxins.

Acknowledgment—We are grateful to Sara Rothman (University of Washington, Seattle, WA) for helpful suggestions concerning isolation of chromosomal DNA.

REFERENCES

Ahnert-Hilger, G., Weller, U., Dauzenroth, M. E., Habermann, E., and Gratzl, M. (1989) *FEBS Lett.* **242,** 245–248

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. E., Smith, J. A., and Struhl, K. (eds) (1987) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York

Bensi, G., Raugi, G., Klefenz, H., and Cortese, R. (1986) *EMBO J.* **4,** 119–126

Betley, M. J., Somers, E., and DasGupta, B. R. (1989) *Biochem. Biophys. Res. Commun.* **162,** 1388–1395

Bittner, M. A., Habig, W. H., and Holz, R. W. (1989) *J. Neurochem.* **53,** 966–968

Black, J. D., and Dolly, J. O. (1986a) *J. Cell Biol.* **103,** 521–534

Black, J. D., and Dolly, J. O. (1986b) *J. Cell Biol.* **103,** 535–544

Boroff, D. A., and Fleck, U. (1966) *J. Bacteriol.* **92,** 1580–1581

Chou, P. Y., and Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.* **47,** 247–254

DasGupta, B. R., and Datta, A. (1988) *Biochimie (Paris)* **70,** 811–817

DasGupta, B. R., and Rasmussen, S. (1984) *Arch. Biochem. Biophys.* **232,** 172–178

Eisel, U., Jarausch, W., Goretzki, K., Henschen, A., Engels, J., Weller, U., Hudel, M., Habermann, E., and Niemann, H. (1986) *EMBO J.* **5,** 2495–2502

Fairweather, N., and Lyness, V. A. (1986) *Nucleic Acids Res.* **14,** 7809–7812

Fujii, N., Oguma, K., Yokosawa, N., Kimura, K., and Tsuzuki, K. (1988) *Appl. Environ. Microbiol.* **54,** 69–73

Gimenez, J., Foley, J., and DasGupta, B. R. (1988) *FASEB J.* **2,** A1750

Graves, M. C., and Rabinowitz, J. C. (1986) *J. Biol. Chem.* **261,** 11409–11415

Kozaki, S., and Sakaguchi, G. (1975) *Infect. Immun.* **11,** 932–936

Krieg, P. A., and Melton, D. A. (1984) *Nucleic Acids Res.* **12,** 7057–7070

Krieglstein, K., Henschen, A., Weller, U., and Habermann, E. (1990) *Eur. J. Biochem.* **188,** 39–45

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Mayer, T., Tamura, T., Falk, M., and Niemann, H. (1988) *J. Biol. Chem.* **263,** 14956–14963

Niemann, H., Andersen-Beckh, B., Binz, T., Eisel, U., Demotz, S., Mayer, T., and Widman, C. (1988) *Zentralbl. Bakteriol. Suppl.* **17,** 29–38

Roa, M., and Boquet, P. (1985) *J. Biol. Chem.* **260,** 6827–6835

Sanger, F., Nicklen, F., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467

Sathyamoorthy, V., and DasGupta, B. R. (1985) *J. Biol. Chem.* **260,** 10461–10466

Sathyamoorthy, V., DasGupta, B. R., and Niece, R. L. (1986) *Fed. Proc.* **45,** 1793

Schmidt, J. J., Sathyamoorthy, V., and DasGupta, B. R. (1984) *Biochem. Biophys. Res. Commun.* **119,** 900–904

Schmidt, J. J., Sathyamoorthy, V., and DasGupta, B. R. (1985) *Arch. Biochem. Biophys.* **238,** 544–548

Simpson, L. L. (1989) in *Botulinum Neurotoxin and Tetanus Toxin* (Simpson, L. L., ed) pp. 153–173, Academic Press, Orlando, FL

Tacket, C. O., and Rogawski, M. A. (1989) in *Botulinum Neurotoxin and Tetanus Toxin* (Simpson, L. L., ed) pp. 351–372, Academic Press, Orlando, FL

Tinoco, I., Borer, P. M., Dengler, B., and Levine, M. D. (1973) *Nature* **264,** 40–41

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene (Amst.)* **33,** 103–119